

МИНОБРНАУКИ РОССИИ

РГУ НЕФТИ И ГАЗА (НИУ) ИМЕНИ И.М. ГУБКИНА

Факультет Автоматики и вычислительной техники
Кафедра Информатики

Оценка комиссии: _____ Рейтинг: _____
Подписи членов комиссии:

_____	Вишневская Е.А.
(подпись)	(фамилия, имя, отчество)
_____	_____
(подпись)	(фамилия, имя, отчество)

(дата)	

КУРСОВАЯ РАБОТА

по дисциплине Средства интеллектуального анализа данных и машинное обучение

на тему Применение средств машинного обучения для интеллектуального анализа данных.

«К ЗАЩИТЕ»

к.ф.м.н, доцент кафедры информатики,
Вишневская Е.А.
(должность, ученая степень; фамилия, и.о.)

(подпись)

(дата)

ВЫПОЛНИЛ:

Студент группы АА-22-08
(номер группы)

Коргин Артём Денисович
(фамилия, имя, отчество)

(подпись)

(дата)

Москва, 20 25

МИНОБРНАУКИ РОССИИ

РГУ НЕФТИ И ГАЗА (НИУ) ИМЕНИ И.М. ГУБКИНА

Факультет Автоматики и вычислительной техники

Кафедра Информатики

ЗАДАНИЕ НА КУРСОВУЮ РАБОТУ

по дисциплине Средства интеллектуального анализа данных и машинное обучение

на тему Применение средств машинного обучения для интеллектуального анализа данных

ДАНО студенту Коргину Артёму Денисовичу группы АА-22-08
(фамилия, имя, отчество в дательном падеже) (номер группы)

Содержание работы:

1. Введение
2. Теоретическая часть
3. Практическая часть
4. Заключение
5. Список используемых источников
6. Приложения

Исходные данные для выполнения работы:

1. Данные из электронных источников

Рекомендуемая литература:

1. Кохави Р., Танг Д., Сюй Я. *Доверительное A/B-тестирование. Практическое руководство по контролируемым экспериментам*. М.: ДМК Пресс, 2021. 298 с.
2. Брюс П., Брюс Э. *Практическая статистика для специалистов Data Science*. СПб.: Питер, 2018. 304 с.

Графическая часть:

1. _____

Требования к представлению результатов:

✓	Электронная версия
	Бумажный вариант и электронный образ документа

Руководитель: к.ф.м.н. Доцент Вишневская Е.А.
(уч.степень) (должность) (подпись) (фамилия, имя, отчество)

Задание принял к исполнению: студент Коргин А.Д.
(подпись) (фамилия, имя, отчество)

СОДЕРЖАНИЕ

Введение	5
Теоретическая часть	8
Этапы проведения А/В-тестирования.....	8
Практическая часть	12
Заключение.....	23
Список используемых источников	25
Приложения.....	26
Приложение 1	26
Приложение 2.....	28
Приложение 3.....	39

ВВЕДЕНИЕ

В современном цифровом маркетинге принятие решений на основе данных становится ключевым фактором успешного развития бизнеса. Одним из наиболее распространённых и эффективных инструментов для оценки влияния изменений в маркетинговых стратегиях является метод АВ-тестирования. Этот метод позволяет экспериментально сравнивать две маркетинговые стратегии для выбора наиболее эффективной, которая сможет конвертировать трафик в продажи (или трафик в другую целевую метрику) максимально эффективно и результативно. Это одна из ключевых концепций, которую должен знать каждый специалист в области Data Science.

Актуальность исследования обусловлена растущей потребностью в глубоких знаниях и практических навыках проведения АВ-тестов, а также в понимании особенностей анализа и интерпретации полученных данных. Несмотря на широкое применение АВ-тестирования, многие специалисты сталкиваются с трудностями при выборе корректных статистических методов, определении размера выборки и интерпретации результатов, что может привести к ошибочным выводам и неэффективным решениям.

Цель данной работы - изучение методики АВ-тестирования на примере оценки двух маркетинговых стратегий, направленных на достижение одной цели. В рамках работы планируется подробно рассмотреть этапы проведения АВ-теста, методы обработки и анализа данных, а также особенности интерпретации результатов.

Объектом исследования выступает процесс проведения АВ-тестирования в маркетинговой практике, а предметом - методические подходы к анализу данных, полученные в ходе эксперимента.

Для достижения поставленной цели в работе решаются следующие задачи:

1. Рассмотреть теоретические основы АВ-тестирования, включая формулировку гипотез и выбор метрик.

2. Изучить методы статистического анализа, применяемые для оценки результатов АВ-теста.
3. Провести практический анализ данных двух маркетинговых стратегий с использованием выбранных методов.

Таким образом, работа направлена на формирование комплексного представления о методике АВ-тестирования, что позволит повысить качество принятия решений в маркетинговой деятельности и обеспечить более эффективное использование экспериментальных данных.

Структура работы:

Данная курсовая работа структурирована следующим образом. В введении - обосновывается актуальность применения А/В-тестирования для анализа эффективности маркетинговых стратегий в современных условиях цифрового маркетинга. Описываются цель, задачи, объект и предмет исследования, а также приводится краткий обзор существующих проблем и мотивация выбора темы.

В теоретической части - подробно рассматриваются основные этапы проведения А/В-тестирования, включая формулировку гипотез, выбор ключевых метрик, методы разделения аудитории и сбора данных. Представлен обзор современных статистических методов анализа результатов экспериментов: t-тест, критерий хи-квадрат, непараметрические тесты, а также расчет эффекта Коэна для оценки практической значимости различий. Особое внимание уделяется вопросам корректности интерпретации результатов и возможным ограничениям выбранных подходов.

В практической части - описывается процесс подготовки и анализа реальных данных двух маркетинговых кампаний. Подробно излагаются этапы предобработки данных, проверки гипотез, расчет и интерпретация статистических показателей, визуализация результатов и анализ воронки конверсии. Для иллюстрации всех этапов приводятся графические материалы.

На основе полученных результатов формулируются выводы о сравнительной эффективности исследуемых стратегий.

В заключении подводятся итоги проведенного исследования, формулируются основные выводы по теоретической и практической частям, оценивается практическая значимость работы и даются рекомендации по дальнейшему использованию и развитию методики А/В-тестирования в маркетинговой практике. Также обозначаются направления для будущих исследований и возможные пути совершенствования аналитических подходов.

Список использованных источников содержит литературу, нормативные документы и электронные ресурсы, на которые опирается работа.

В приложениях представлены листинги программного кода.

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

A/B-тестирование - это фундаментальный метод в Data Science и машинном обучении, который позволяет сравнивать две версии продукта, маркетинговой стратегии или иной бизнес-гипотезы для принятия решений на основе объективных данных. Данный метод широко применяется для оценки влияния изменений на поведение пользователей и ключевые метрики, минимизируя риски при внедрении нововведений.

Этапы проведения A/B-тестирования

1. Определение цели

Первый и ключевой этап - чёткое формулирование цели теста. Цель должна отражать, что именно вы хотите улучшить или проверить. Например, «увеличение конверсии на сайте» или «повышение среднего чека». Чётко сформулированная цель позволяет выбрать релевантные метрики и корректно интерпретировать результаты.[1]

2. Формулировка гипотез

На этом этапе формулируются две гипотезы:

- *Нулевая гипотеза (H_0):* предполагает отсутствие значимых различий между контрольной и тестовой группами. Формально:

$$H_0: \mu_A = \mu_B \quad (1.1)$$

где μ_A и μ_B - средние значения метрики в группах A и B соответственно.

- *Альтернативная гипотеза (H_1):* предполагает наличие статистически значимого различия:

$$H_1: \mu_A \neq \mu_B \quad (1.2)$$

Гипотезы должны быть проверяемыми, конкретными и связаны с целью теста.

3. Разделение аудитории

Для исключения систематических ошибок и смещений пользователей случайным образом распределяют на две группы:

- *Контрольная группа (A)*: получает текущий вариант продукта или стратегии без изменений.
- *Тестовая группа (B)*: взаимодействует с новой версией (например, изменённый дизайн кнопки).

Случайное распределение обеспечивает равенство условий и минимизирует влияние внешних факторов.

4. Внедрение изменений

Изменения внедряются только в тестовой группе, при этом контрольная группа остаётся без изменений. Важно, чтобы тест проходил одновременно для обеих групп, чтобы исключить влияние временных факторов (сезонность, акции и т.п.).

5. Сбор данных

В ходе теста собираются данные по заранее выбранным метрикам, например:

- Конверсия (CR) - доля пользователей, совершивших целевое действие:

$$CR = \frac{\text{Количество целевых действий}}{\text{Общее число пользователей}} \times 100\% \quad (1.3)$$

- Средний чек, количество кликов, время на сайте и другие показатели.

При сборе данных важно исключить смещения, например, из-за технических сбоев или неравномерного распределения трафика.

6. Анализ результатов

Для оценки статистической значимости различий применяются различные методы:

- *t-тест для независимых выборок* - используется при нормальном распределении метрики. Рассчитывается t-статистика:

$$t = \frac{\overline{X_A} - \overline{X_B}}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \quad (1.4)$$

где $\overline{X_A}$, $\overline{X_B}$ - средние значения, S_A^2, S_B^2 - дисперсии, n_A, n_B - размеры выборок.[2]

- *Эффект Козна* - метод оценки размера эффекта.

Эффект Коэна - это одна из наиболее распространённых и важных статистических мер, используемых для количественной оценки величины различий между двумя группами. В отличие от классических тестов значимости (например, t-теста), которые отвечают на вопрос, существует ли статистически значимое различие, эффект Коэна показывает, насколько это различие велико с практической точки зрения. Это особенно важно, поскольку статистическая значимость может быть достигнута при очень больших выборках даже при незначительных, по сути, различиях.

Определение и смысл эффекта Коэна

Эффект Коэна (обозначается как d) измеряет стандартизированную разницу между средними значениями двух групп, выраженную в единицах стандартного отклонения. Его можно интерпретировать как количество стандартных отклонений, на которое среднее значение одной группы отличается от другой. Формула для вычисления эффекта Коэна:

$$d = \frac{\overline{X}_A - \overline{X}_B}{S_p} \quad (1.5)$$

где:

- $\overline{X}_A, \overline{X}_B$ - средние значения двух групп,
- S_p - объединённое стандартное отклонение.

Объединённое стандартное отклонение рассчитывается по формуле:

$$S_p = \sqrt{\frac{(n_A - 1)S_A + (n_B - 1)S_B}{n_A + n_B - 2}} \quad (1.6)$$

где:

- S_A, S_B - стандартные отклонения групп,
- n_A, n_B - размеры выборок.

Интерпретация величины эффекта Коэна

Джейкоб Коэн предложил следующие ориентиры для интерпретации значения d :

- $d = 0.2$ - малый эффект (небольшая, но заметная разница),

- $d = 0.5$ - средний эффект (умеренная разница),
- $d = 0.8$ и выше - большой эффект (существенная разница).

Эти пороги являются рекомендациями, а не строгими правилами, и их следует учитывать в контексте конкретной области исследования.[3]

- *Критерий хи-квадрат* - применяется для анализа категориальных данных (например, распределения кликов по вариантам).
- *Непараметрические тесты (например, Манна-Уитни)* - используются при отсутствии нормальности распределения.
- *Регрессионные модели* - позволяют учитывать дополнительные факторы (возраст, регион, устройство) и оценивать их влияние на результат.

Для проверки нормальности данных применяют тест Шапиро-Уилка. Если $p\text{-value} > 0.05$, данные считаются нормально распределёнными.[2]

7. Принятие решения

На основании рассчитанного $p\text{-value}$ принимается решение:

- Если $p < \alpha$ (обычно $\alpha = 0.05$), нулевая гипотеза отвергается, и считается, что изменения оказали значимое влияние.
- Если $p \geq \alpha$, статистически значимых различий нет, и изменения не считаются эффективными.

8. Внедрение изменений

Если тест показал эффективность новой версии, изменения масштабируются на всю аудиторию. При отсутствии значимых улучшений либо проводится повторное тестирование с другими гипотезами.

Итог

А/В-тестирование - это системный и статистически обоснованный подход, который позволяет снижать риски при внедрении изменений, принимать решения на основе данных, а не интуиции, и оптимизировать продукты и модели машинного обучения. Правильное планирование, сбор и анализ данных в сочетании с учётом внешних факторов обеспечивают надёжность и воспроизводимость результатов.

ПРАКТИЧЕСКАЯ ЧАСТЬ

1. Определение цели

Целью данного А/В-теста является изучение и демонстрация методики анализа данных с помощью А/В-тестирования на примере сравнения двух маркетинговых стратегий. В частности, задача - определить, влияет ли изменение маркетингового подхода на ключевую бизнес-метрику - количество покупок

2. Формулировка гипотез

- *Нулевая гипотеза (H_0):* Между двумя маркетинговыми стратегиями нет статистически значимых различий по количеству покупок.

$$H_0: \mu_A = \mu_B \quad (2.1)$$

- *Альтернативная гипотеза (H_1):* Изменение маркетинговой стратегии приводит к статистически значимым изменениям в количестве покупок.

$$H_1: \mu_A \neq \mu_B \quad (2.2)$$

3. Разделение аудитории

Для проведения теста аудитория была случайным образом разделена на две группы:

- *Контрольная группа (A):* пользователи, подвергшиеся действующей маркетинговой стратегии.
- *Тестовая группа (B):* пользователи, подвергшиеся изменённой маркетинговой стратегии.

Данные по каждой группе собраны за равные по длительности периоды и включают показатели ежедневных покупок и других ключевых метрик.

4. Обработка данных

- Ошибки в названиях столбцов были исправлены.
- Пропуски в данных были заполнены медианными значениями для сохранения статистической устойчивости.
- Данные были объединены в один датасет и проверены на одинаковое количество наблюдений, содержащееся в обеих кампаниях.

- Проверена нормальность распределения метрики "Покупки" с помощью теста Шапиро-Уилка. Результаты показали, что данные близки к нормальному распределению ($p\text{-value} > 0.05$), что позволяет использовать параметрические методы анализа.

Листинг кода представлен в Приложении 1.

5. Анализ результатов

Для начала А/В-тестирования я проанализирую взаимосвязь между количеством показов из обеих кампаний и затраченными на них средствами:

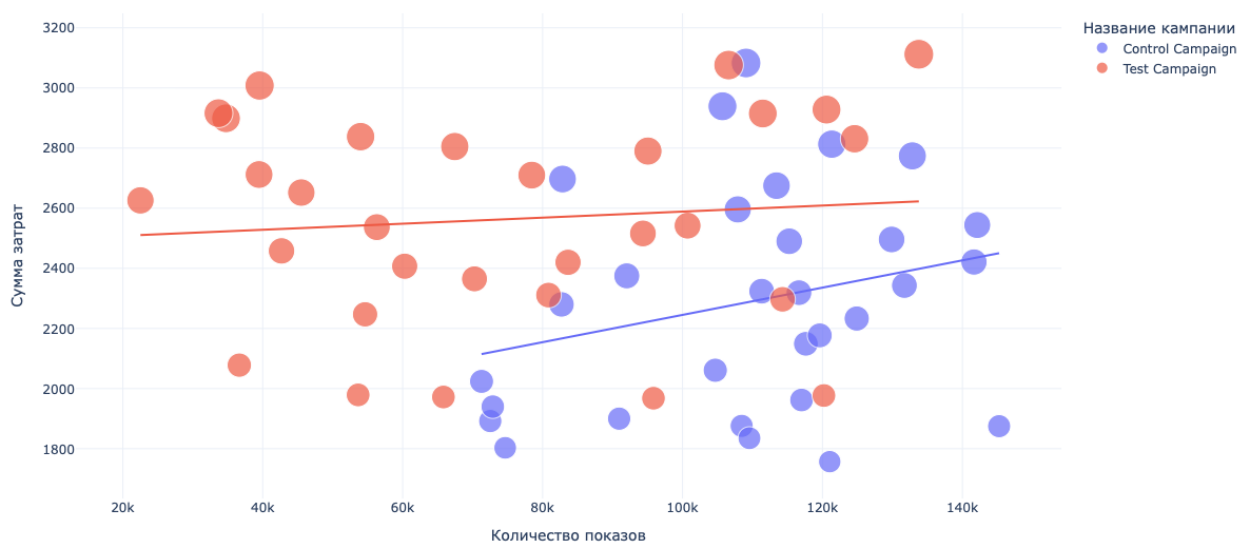


Рисунок 2.1. Лучшая маркетинговая стратегия

Согласно затраченной сумме, контрольная кампания обеспечила больше показов Рисунок 2.1. Теперь проанализируем количество поисковых запросов на сайте для обеих кампаний:

- 1) $p\text{-value} = 0.2540 (> 0.05)$: Различия не значимы.
- 2) $d = -0.30$:
 - Test Campaign слегка лидирует, но разница несущественна.
 - Эффект малый, статистической значимости нет.

Тестовая кампания привела к большему количеству поисков на сайте Рисунок 2.2.

Контрольная vs Тестовая: Поисковые запросы

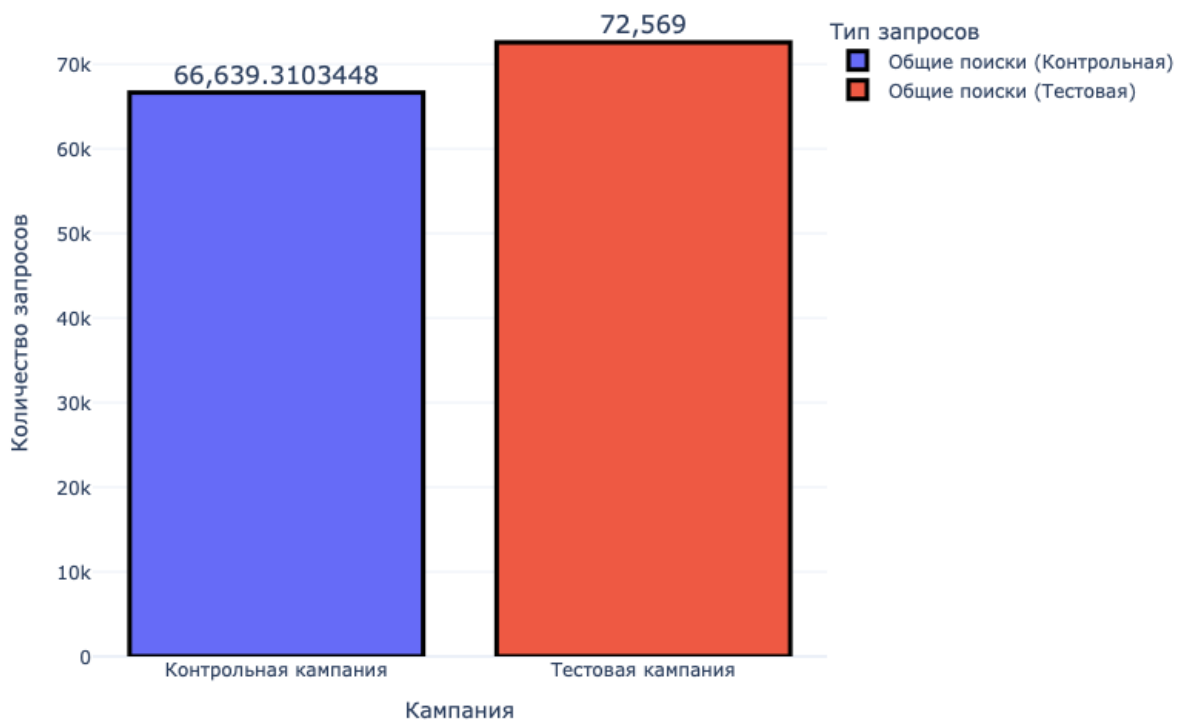


Рисунок 2.2. Общее количество поисковых запросов

Теперь проанализируем количество кликов на сайте для обеих кампаний:

Контрольная vs Тестовая: Клики на сайте

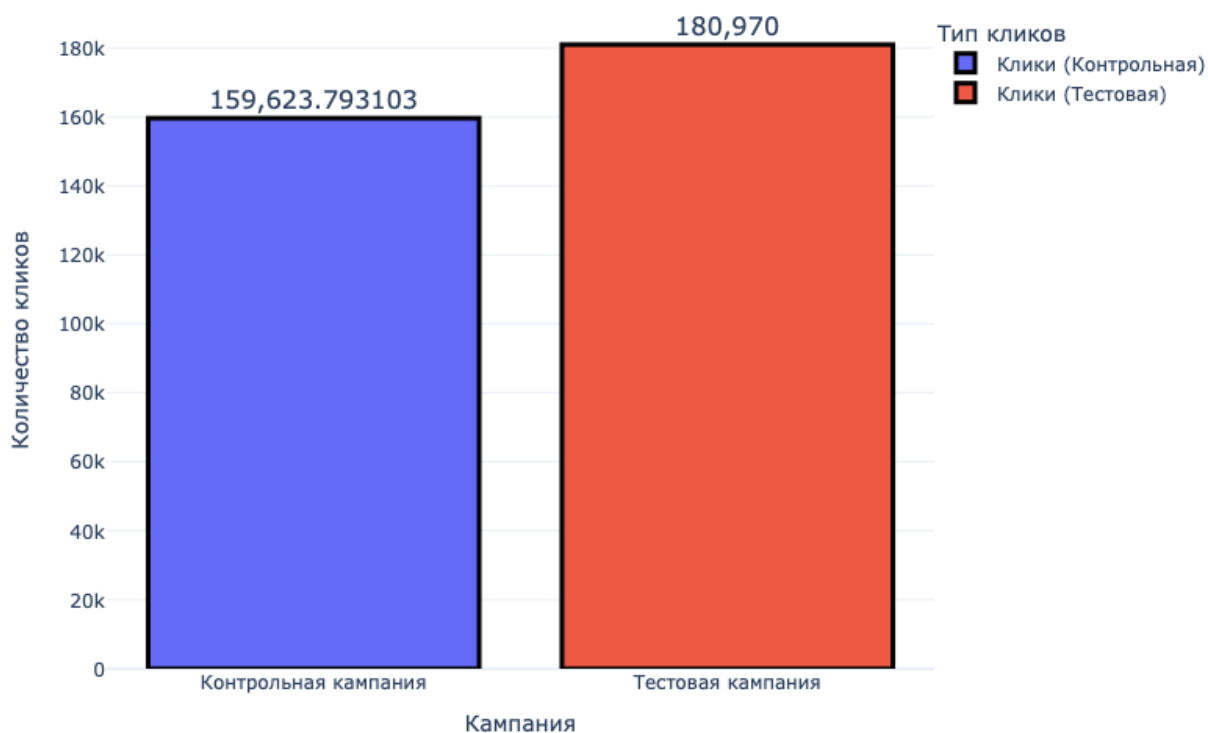


Рисунок 2.3. Общее количество кликов на сайте

1) $p\text{-value} = 0.1141 (> 0.05)$: Различия не значимы.

2) $d = -0.42$:

- Test Campaign получила немного больше кликов, но из-за высокого $p\text{-value}$ это может быть случайностью.
- Эффект средний, но не статистически подтвержденный.

Тестовая кампания лидирует по количеству кликов на сайте Рисунок 2.3.

Теперь проанализируем, какой объём контента был просмотрен после перехода на сайт в обеих кампаниях:

Контрольная vs Тестовая: Просмотры контента

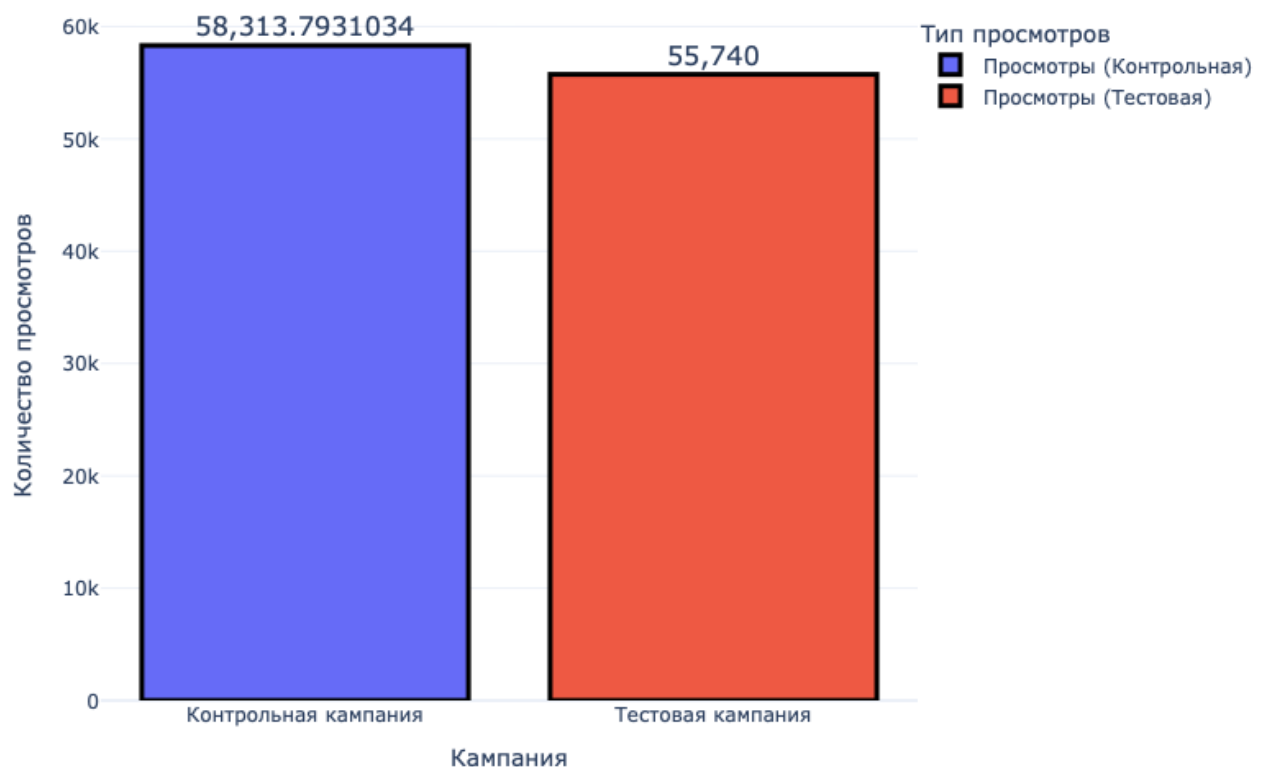


Рисунок 2.4. Общее количество просмотров контента

1) $p\text{-value} = 0.6300 (> 0.05)$: Различия не значимы.

2) $d = 0.13$:

- Почти нулевой эффект - кампании одинаковы по вовлеченности в контент.

Аудитория контрольной кампании просмотрела больше контента, чем тестовой

Рисунок 2.4. Хотя разница невелика, учитывая, что кликов на сайте в

контрольной кампании было меньше, её вовлеченность на сайте выше, чем у тестовой кампании.

Проанализируем количество товаров, добавленных в корзину, для обеих кампаний:

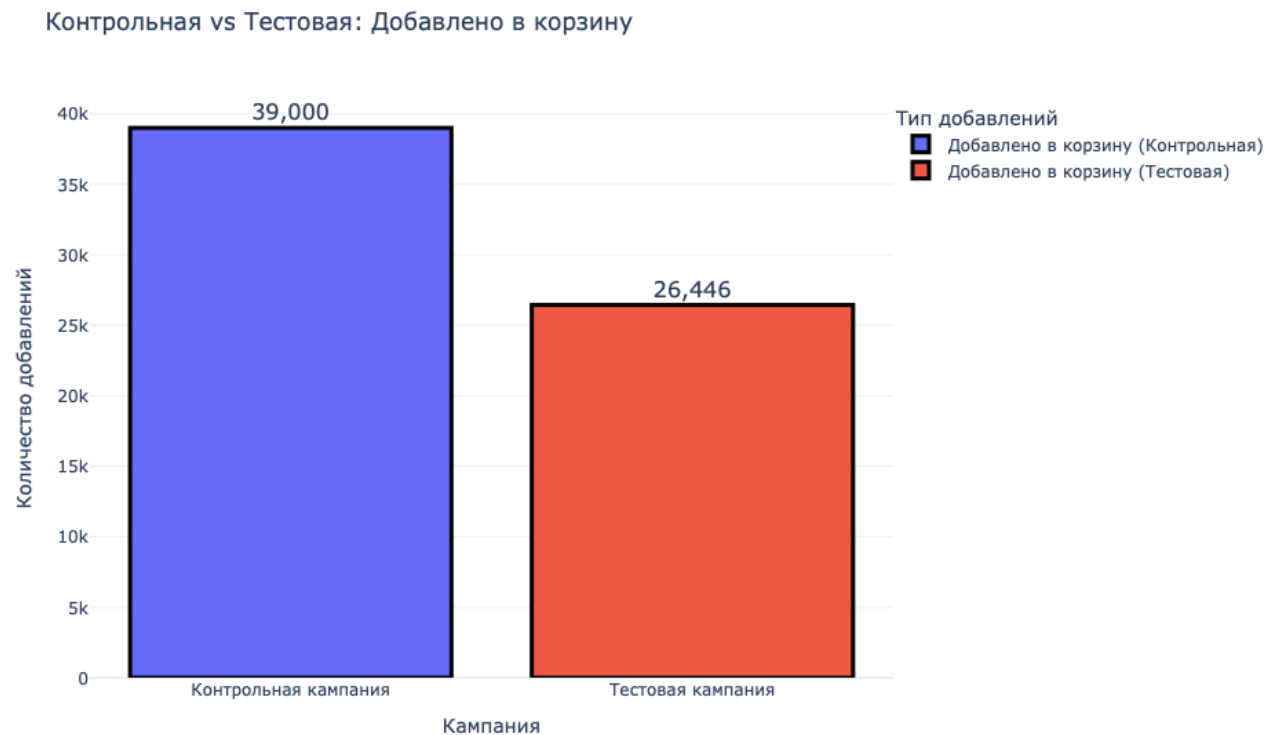


Рисунок 2.5. Общее количество товаров, добавленных в корзину

1) $p\text{-value} = 0.0001$: Различия значимы.

2) $d = 1.14$:

- Control Campaign значительно лучше конвертирует в корзину.
- Очень большой эффект - ключевое преимущество Control Campaign.

Несмотря на низкое количество кликов на сайте, в контрольной кампании в корзину было добавлено больше товаров Рисунок 2.5.

Проанализируем затраченные средства на обеих кампаниях:

1) $p\text{-value} = 0.0043 (< 0.05)$: Различия в затратах статистически значимы.

2) Эффект Коэна (d) = -0.78:

- Отрицательное значение означает, что Test Campaign была дороже Control Campaign.

- Размер эффекта большой, что указывает на существенную разницу в бюджетах.

Контрольная vs Тестовая: Сумма затрат

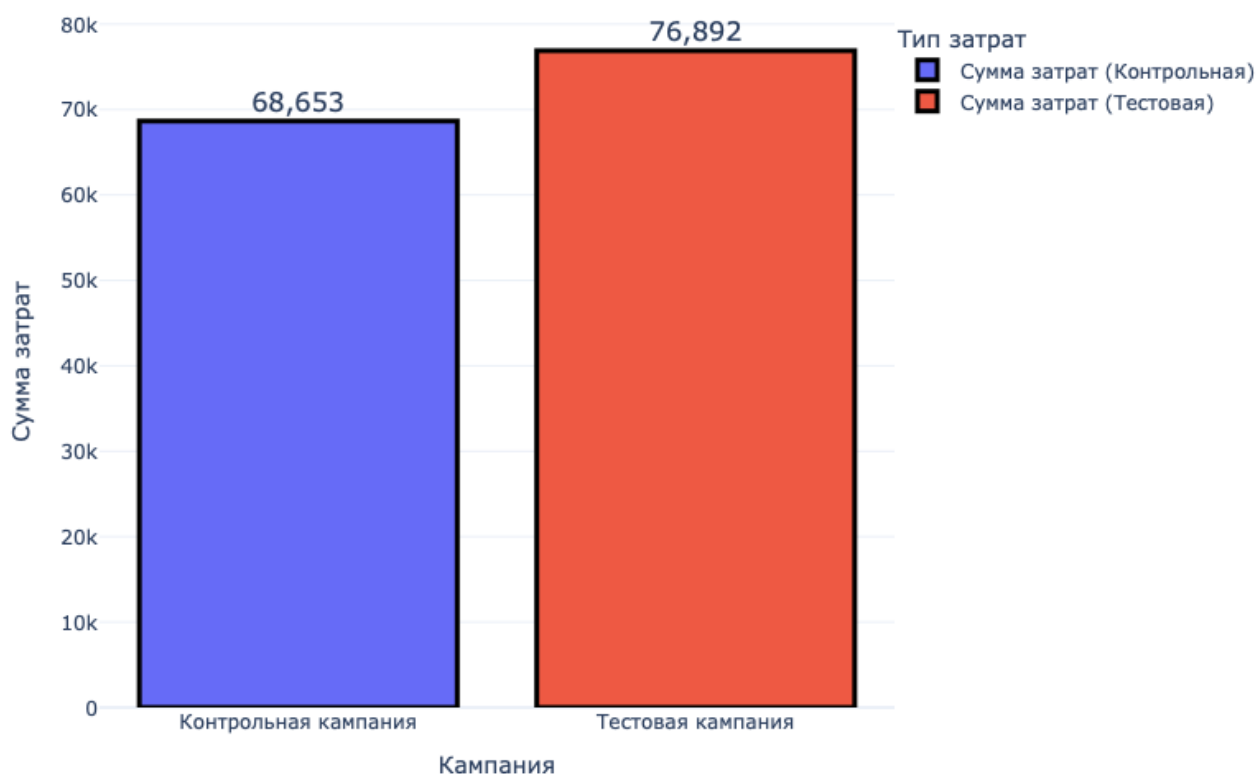


Рисунок 2.6. Общее количество затраченных средств

Затраты на тестовую кампанию выше, чем на контрольную Рисунок 2.6.

Однако, как видно из данных, контрольная кампания обеспечила больше просмотров контента и товаров в корзине, что делает её более эффективной по сравнению с тестовой.

Проанализируем покупки, совершённые в обеих кампаниях:

- 1) $p\text{-value} = 0.9756$: Различия не значимы.
- 2) $d = 0.01$:
 - Нулевой эффект — конечные продажи одинаковы, несмотря на различия на предыдущих этапах.

Контрольная vs Тестовая: Покупки

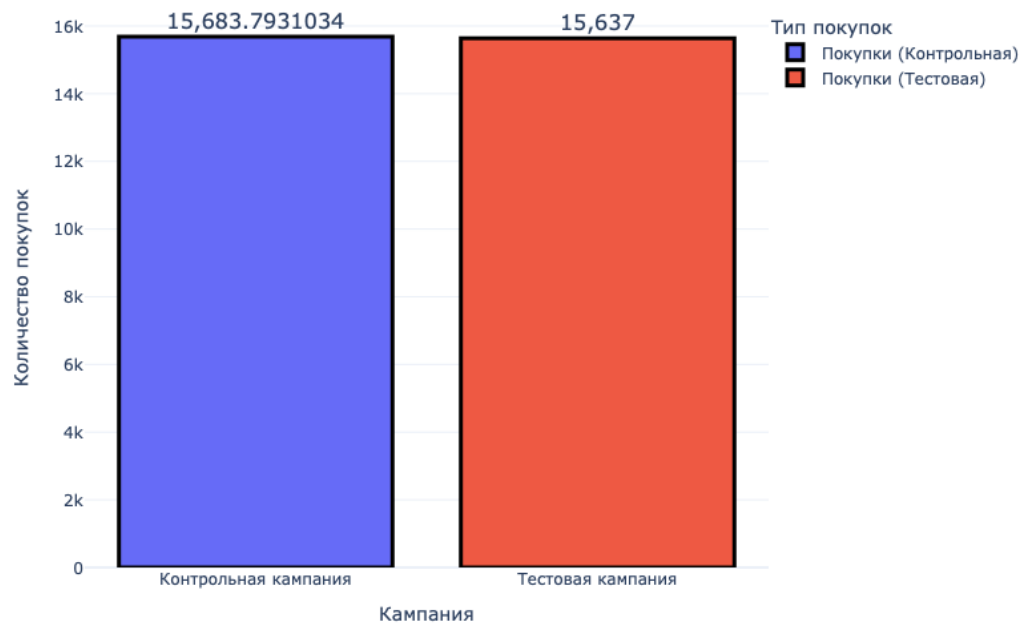


Рисунок 2.7. Общее количество совершенных покупок

Разница в покупках между рекламными кампаниями составляет около 1%. Поскольку контрольная кампания привела к большему количеству продаж при меньшем бюджете на маркетинг - контрольная кампания одерживает победу! Листинг кода представлен в Приложении 2.

6. Анализ метрик для определения конверсии.

Проанализируем метрики, чтобы определить, какая рекламная кампания обеспечивает более высокую конверсию. Сначала я изучу взаимосвязь между количеством кликов на сайте и просмотрами контента в обеих кампаниях:

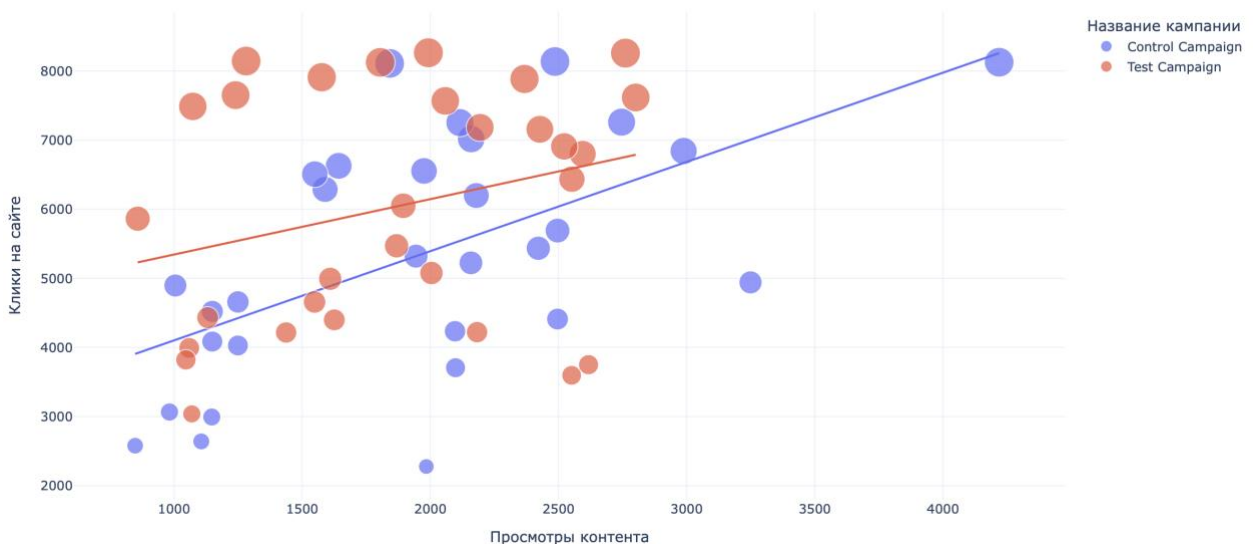


Рисунок 2.8. Просмотры контента относительно кликов на сайте

Количество кликов на сайте выше в тестовой кампании, но вовлеченность (конверсия) с этих кликов выше в контрольной кампании Рисунок 2.8.

Победитель - контрольная кампания!

Дальше я проанализирую взаимосвязь между объёмом просмотренного контента и количеством товаров, добавленных в корзину, для обеих кампаний:



Рисунок 2.9. Добавление в корзину относительно просмотров контента

В этой взаимосвязи тоже побеждает контрольная кампания.

Проанализируем взаимосвязь между количеством товаров, добавленных в корзину, и объёмом продаж для обеих кампаний:

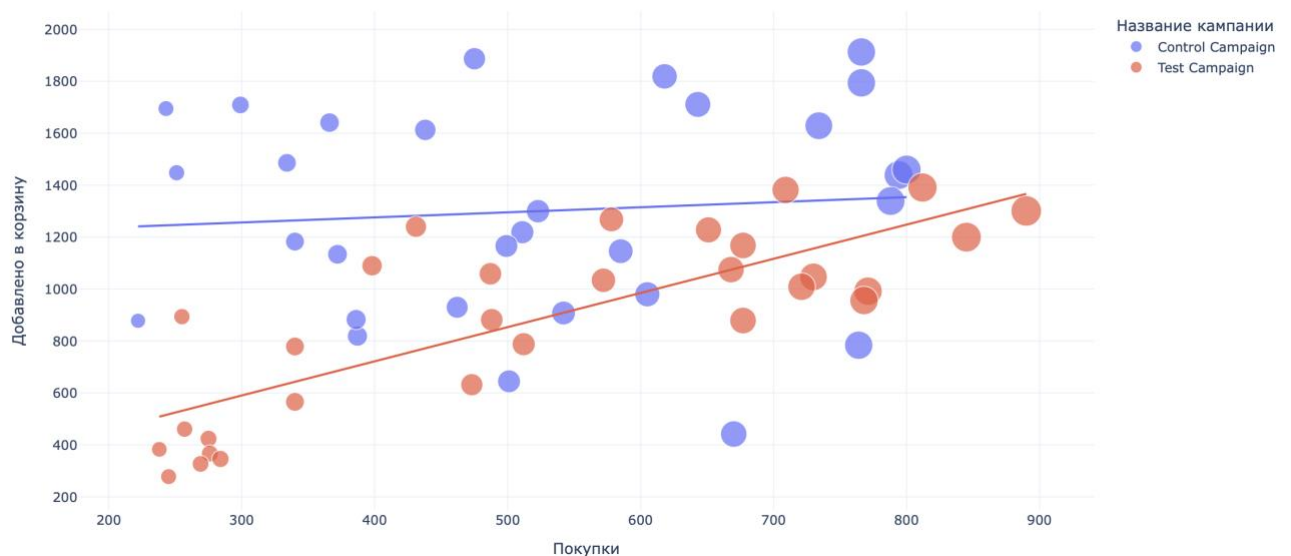


Рисунок 2.10. Покупки относительно добавления в корзину

Хотя контрольная кампания привела к большему количеству продаж и товаров в корзине, конверсия тестовой кампании выше.

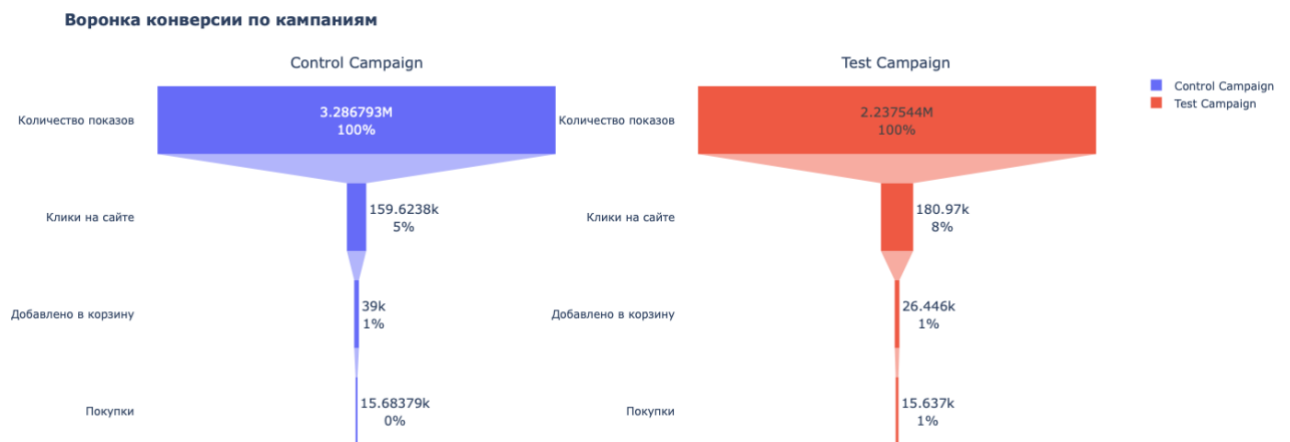


Рисунок 2.11. Воронка конверсий по кампаниям

Проанализируем воронку конверсии для Control Campaign и Test Campaign:

1) Этапы воронки и ключевые метрики:

- Количество показов:
 - *Control Campaign*: 3,286,793 (100%)
 - *Test Campaign*: 2,237,544 (100%)
 - *Интерпретация*: Test Campaign охватила меньшую аудиторию, но это могло быть сознательным решением (например, таргетинг на узкую группу).
- Клики на сайте:
 - *Control Campaign*: 159,623 (5% от показов)
 - *Test Campaign*: 180,970 (8% от показов)
 - *Интерпретация*: Test Campaign эффективнее привлекает пользователей к переходу на сайт. Возможно, из-за более релевантного контента или креативов.
- Добавлено в корзину:
 - *Control Campaign*: 39,039 (1% от кликов)
 - *Test Campaign*: 26,446 (1% от кликов)

- *Интерпретация:* Несмотря на меньшее число кликов, Control Campaign лучше конвертирует их добавления в корзину. Возможно, из-за удобства интерфейса или ценообразования.
- Покупки:
 - *Control Campaign:* 15,683 (0.4% от добавлений в корзину)
 - *Test Campaign:* 15,637 (0.6% от добавлений в корзину)
 - *Интерпретация:* Test Campaign показывает чуть лучшую конверсию в покупке, но абсолютные значения почти одинаковы. Возможно, из-за различий в условиях доставки, акциях или UX финального этапа.

2) Ключевые выводы:

- Test Campaign:
 - Лучшая конверсия на этапе кликов (+3% к Control Campaign).
 - Хуже удерживает пользователей на этапе корзины.
 - Чуть выше конверсия в покупке, но незначительно.
- Control Campaign:
 - Большой охват аудитории.
 - Эффективнее конвертирует клики в добавления в корзину.
 - Почти идентичный результат по покупкам, несмотря на изначально большую аудиторию.

3) Рекомендации:

- Для Test Campaign:
 - Улучшить этап перехода из корзины в покупку (например, упростить процесс оплаты).
 - Проанализировать причины низкой конверсии из кликов в корзину (возможно, проблемы с ассортиментом или ценами).
- Для Control Campaign:
 - Повысить конверсию из показов в клики (оптимизировать рекламные креативы).

- Увеличить конверсию из корзины в покупки (например, добавить напоминания о брошенной корзине).

4) Визуализация:

График показывает, что Test Campaign эффективнее на ранних этапах, но проигрывает на финальных. Control Campaign демонстрирует стабильность, но требует оптимизации для увеличения конверсии из показов в клики. Обе кампании имеют потенциал для улучшения, особенно на этапе завершения покупки.

Листинг кода представлен в Приложении 3.

7. Принятие решения

По результатам проведённых А/В-тестов, контрольная кампания показала:

- *Более высокие продажи и вовлечённость:* Посетители чаще просматривали товары, добавляли их в корзину и совершали покупки.
- *Широкий охват:* Эффективна для продвижения ассортимента товаров массовой аудитории.

Тестовая кампания выделилась:

- *Высокой конверсией:* Лучшее соотношение добавлений в корзину к покупкам для целевых товаров.
- *Точечным воздействием:* Подходит для продвижения специфических продуктов узкой аудитории.

Рекомендация:

Используйте тестовую кампанию для точечных промо-акций, а контрольную - для масштабных распродаж с широким ассортиментом.

ЗАКЛЮЧЕНИЕ

В ходе выполнения курсовой работы было проведено комплексное исследование методики А/В-тестирования на примере сравнения двух маркетинговых стратегий. В теоретической части подробно рассмотрены основные этапы проведения А/В-теста, включая формулировку гипотез, выбор метрик, методы статистического анализа и интерпретацию результатов. Особое внимание уделено применению классического статистического теста (t-тест), а также оценке размера эффекта с помощью эффекта Коэна, что обеспечивает более глубокое понимание значимости выявленных различий.

Практическая часть включала сбор и предобработку данных по контрольной и тестовой группам, проведение статистического анализа и визуализацию результатов. Анализ воронки конверсии показал, что тестовая маркетинговая стратегия эффективнее привлекает пользователей на ранних этапах (показы → клики), однако уступает контрольной группе в удержании пользователей на этапе добавления товаров в корзину. Итоговые показатели по количеству покупок оказались практически равны, что свидетельствует о необходимости дальнейшей оптимизации финальных этапов воронки для обеих кампаний.

Цель работы - изучение методики А/В-тестирования на конкретном примере - была успешно достигнута. Решены все поставленные задачи: рассмотрены теоретические основы А/В-тестирования, сформулированы и проверены гипотезы, изучены и проведены методы статистического анализа, сделаны выводы о практической эффективности маркетинговых стратегий.

Практическая значимость работы заключается в демонстрации системного подхода к проведению А/В-тестирования и анализу результатов, что позволяет принимать обоснованные решения на основе данных и минимизировать риски при внедрении изменений. Рекомендации по улучшению конверсии на различных этапах воронки могут быть использованы для повышения эффективности маркетинговых кампаний в реальных условиях.

В дальнейшем целесообразно расширить исследование за счёт включения сегментационного анализа пользователей с применением методов машинного обучения, а также проведения многовариантного тестирования (A/B/n) для более комплексной оптимизации маркетинговых стратегий. Кроме того, планируется изучение влияния внешних факторов (сезонность, поведение конкурентов) на результаты A/B-тестов для повышения точности и надежности выводов.

Таким образом, выполненная работа вносит вклад в развитие практических навыков анализа данных и применения современных методов A/B-тестирования в маркетинге и Data Science.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Кохави Р., Танг Д., Сюй Я. *Доверительное A/B-тестирование. Практическое руководство по контролируемым экспериментам*. М.: ДМК Пресс, 2021. 298 с.
2. Брюс П., Брюс Э. *Практическая статистика для специалистов Data Science*. СПб.: Питер, 2018. 304 с.
3. Джеймс Г., Уиттон Д., Хасты Т. *Введение в статистическое обучение с примерами на языке R*. М.: ДИАЛЕКТИКА, 2020. 440 с.
4. Документация SciPy: [Электронный ресурс].
URL: <https://docs.scipy.org/> (дата обращения: 15.05.2025).

ПРИЛОЖЕНИЯ

Приложение 1

Обработка данных:

Чтение данных

```
control_data = pd.read_csv("control_group.csv", sep = ";")
```

```
test_data = pd.read_csv("test_group.csv", sep = ";")
```

Подготовка данных

```
control_data.columns = ["Название кампании", "Дата", "Сумма затрат",  
                        "Количество показов", "Охват", "Клики на сайте",  
                        "Поисковые запросы", "Просмотры контента", "Добавлено в  
корзину",  
                        "Покупки"]
```

```
test_data.columns = ["Название кампании", "Дата", "Сумма затрат",  
                    "Количество показов", "Охват", "Клики на сайте",  
                    "Поисковые запросы", "Просмотры контента", "Добавлено в  
корзину",  
                    "Покупки"]
```

Проверка данных

```
control_data.isnull().sum()
```

```
test_data.isnull().sum()
```

Пропущенные значения

```
control_data["Количество показов"].fillna(value=control_data["Количество  
показов"].mean(), inplace=True)
```

```
control_data["Охват"].fillna(value=control_data["Охват"].mean(), inplace=True)
```

```
control_data["Клики на сайте"].fillna(value=control_data["Клики на  
сайте"].mean(), inplace=True)
```

```
control_data["Поисковые запросы"].fillna(value=control_data["Поисковые  
запросы"].mean(), inplace=True)
```

```
control_data["Просмотры контента"].fillna(value=control_data["Просмотры
контента"].mean(), inplace=True)
control_data["Добавлено в корзину"].fillna(value=control_data["Добавлено в
корзину"].mean(), inplace=True)
control_data["Покупки"].fillna(value=control_data["Покупки"].mean(),
inplace=True)
# Объединение наборов данных
ab_data = control_data.merge(test_data, how="outer").sort_values(["Дата"])
ab_data = ab_data.reset_index(drop=True)
# Проверка данных на нормальность распределения
stat_control, p_value_control = stats.shapiro(control_data['Покупки'])
stat_test, p_value_test = stats.shapiro(test_data['Покупки'])
# Проверка датасета на одинаковое содержания наблюдений в данных
ab_data["Название кампании"].value_counts()
```

Приложение 2

Анализ результатов:

```
figure = px.scatter(data_frame=ab_data,
                    x="Количество показов",
                    y="Сумма затрат",
                    size="Сумма затрат",
                    color="Название кампании",
                    trendline="ols")
figure.update_layout(
    width=1200,
    height=600
)
figure.show()
counts = [sum(control_data["Поисковые запросы"]), sum(test_data["Поисковые
запросы"])]
colors = ['#636EFA', '#EF553B']
fig = go.Figure()
fig.add_trace(go.Bar(
    x=["Контрольная кампания"],
    y=[counts[0]],
    name="Общие поиски (Контрольная)",
    marker_color=colors[0],
    marker_line=dict(color='black', width=3))
)
fig.add_trace(go.Bar(
    x=["Тестовая кампания"],
    y=[counts[1]],
    name="Общие поиски (Тестовая)",
    marker_color=colors[1],
```

```

        marker_line=dict(color='black', width=3))
    )
    fig.update_layout(
        title_text='Контрольная vs Тестовая: Поисковые запросы',
        xaxis_title="Кампания",
        yaxis_title="Количество запросов",
        width=800,
        height=600,
        showlegend=True,
        legend=dict(
            title="Тип запросов",
            x=1.0,
            y=1.0,
            bgcolor='rgba(255, 255, 255, 0.5)'
        ),
        uniformtext_minsize=12
    )
    fig.update_traces(
        texttemplate='%{y:,.}',
        textposition='outside',
        textfont_size=16,
        textangle=0
    )
    fig.show()

# #### Статистический анализ [4]
def calculate_statistical_significance(control, test, metric_name):
    """Расчет t-теста и размера эффекта"""
    t_stat, p_value = stats.ttest_ind(control, test, equal_var=False)
    pooled_std = np.sqrt((np.var(control) + np.var(test)) / 2)

```

```

cohens_d = (np.mean(control) - np.mean(test)) / pooled_std
print(f'\nМетрика: {metric_name}')
print(f'p-value: {p_value:.4f}')
print(f'Эффект Коэна: {cohens_d:.2f}')
# Проверка значимости
calculate_statistical_significance(
    control_data['Поисковые запросы'],
    test_data['Поисковые запросы'],
    'Поисковые запросы'
)
# Данные для графика
counts = [sum(control_data["Клики на сайте"]), sum(test_data["Клики на сайте"])]
colors = ['#636EFA', '#EF553B']
fig = go.Figure()
# Добавление столбцов для каждой кампании
fig.add_trace(go.Bar(
    x=["Контрольная кампания"],
    y=[counts[0]],
    name="Клики (Контрольная)",
    marker_color=colors[0],
    marker_line=dict(color='black', width=3)
))
fig.add_trace(go.Bar(
    x=["Тестовая кампания"],
    y=[counts[1]],
    name="Клики (Тестовая)",
    marker_color=colors[1],
    marker_line=dict(color='black', width=3)
))

```

```

# Настройка макета
fig.update_layout(
    title_text='Контрольная vs Тестовая: Клики на сайте',
    xaxis_title="Кампания",
    yaxis_title="Количество кликов",
    width=800,
    height=600,
    showlegend=True,
    legend=dict(
        title="Тип кликов",
        x=1.0,
        y=1.0,
        bgcolor='rgba(255, 255, 255, 0.5)'
    ),
    uniformtext_minsize=12
)

# Настройка отображения значений
fig.update_traces(
    texttemplate='%{y:,.}',
    textposition='outside',
    textfont_size=16,
    textangle=0
)

fig.show()

# Проверка значимости
calculate_statistical_significance(
    control_data['Клики на сайте'],
    test_data['Клики на сайте'],
    'Клики на сайте'
)

```

```

)
# Данные для графика
counts = [sum(control_data["Просмотры контента"]), sum(test_data["Просмотры
контента"])]
colors = ['#636EFA', '#EF553B']
fig = go.Figure()
# Добавление столбцов для каждой кампании
fig.add_trace(go.Bar(
    x=["Контрольная кампания"],
    y=[counts[0]],
    name="Просмотры (Контрольная)",
    marker_color=colors[0],
    marker_line=dict(color='black', width=3)
))
fig.add_trace(go.Bar(
    x=["Тестовая кампания"],
    y=[counts[1]],
    name="Просмотры (Тестовая)",
    marker_color=colors[1],
    marker_line=dict(color='black', width=3)
))
# Настройка макета
fig.update_layout(
    title_text='Контрольная vs Тестовая: Просмотры контента',
    xaxis_title="Кампания",
    yaxis_title="Количество просмотров",
    width=800,
    height=600,
    showlegend=True,

```



```

legend=dict(
    title="Тип просмотров",
    x=1.0,
    y=1.0,
    bgcolor='rgba(255, 255, 255, 0.5)'
),
uniformtext_minsize=12
)
# Настройка отображения значений
fig.update_traces(
    texttemplate='%{y:,.}',
    textposition='outside',
    textfont_size=16,
    textangle=0
)
fig.show()
# Проверка значимости
calculate_statistical_significance(
    control_data['Просмотры контента'],
    test_data['Просмотры контента'],
    'Просмотры контента'
)
# Данные для графика
counts = [sum(control_data["Добавлено в корзину"]),
          sum(test_data["Добавлено в корзину"])]
colors = ['#636EFA', '#EF553B']
fig = go.Figure()
# Добавление столбцов для каждой кампании
fig.add_trace(go.Bar(

```

```

x=["Контрольная кампания"],
y=[counts[0]],
name="Добавлено в корзину (Контрольная)",
marker_color=colors[0],
marker_line=dict(color='black', width=3)
))
fig.add_trace(go.Bar(
    x=["Тестовая кампания"],
    y=[counts[1]],
    name="Добавлено в корзину (Тестовая)",
    marker_color=colors[1],
    marker_line=dict(color='black', width=3))
)
# Настройка макета
fig.update_layout(
    title_text='Контрольная vs Тестовая: Добавлено в корзину',
    xaxis_title="Кампания",
    yaxis_title="Количество добавлений",
    width=950,
    height=600,
    showlegend=True,
    legend=dict(
        title="Тип добавлений",
        x=1.0,
        y=1.0,
        bgcolor='rgba(255, 255, 255, 0.5)'
    ),
    uniformtext_minsize=12
)

```

```

# Настройка отображения значений
fig.update_traces(
    texttemplate='%{y:,.}',
    textposition='outside',
    textfont_size=16,
    textangle=0
)
fig.show()

# Проверка значимости
calculate_statistical_significance(
    control_data['Добавлено в корзину'],
    test_data['Добавлено в корзину'],
    'Добавлено в корзину'
)

# Данные для графика
counts = [sum(control_data["Сумма затрат"]), sum(test_data["Сумма затрат"])]
colors = ['#636EFA', '#EF553B']
fig = go.Figure()

# Добавление столбцов для каждой кампании
fig.add_trace(go.Bar(
    x=["Контрольная кампания"],
    y=[counts[0]],
    name="Сумма затрат (Контрольная)",
    marker_color=colors[0],
    marker_line=dict(color='black', width=3)
))

fig.add_trace(go.Bar(
    x=["Тестовая кампания"],
    y=[counts[1]],

```

```

    name="Сумма затрат (Тестовая)",
    marker_color=colors[1],
    marker_line=dict(color='black', width=3)
))
# Настройка макета
fig.update_layout(
    title_text='Контрольная vs Тестовая: Сумма затрат',
    xaxis_title="Кампания",
    yaxis_title="Сумма затрат",
    width=800,
    height=600,
    showlegend=True,
    legend=dict(
        title="Тип затрат",
        x=1.0,
        y=1.0,
        bgcolor='rgba(255, 255, 255, 0.5)'
    ),
    uniformtext_minsize=12
)
# Настройка отображения значений
fig.update_traces(
    texttemplate='%{y:,.0f}',
    textposition='outside',
    textfont_size=16,
    textangle=0
)
fig.show()
# Проверка значимости

```

```

calculate_statistical_significance(
    control_data['Сумма затрат'],
    test_data['Сумма затрат'],
    'Сумма затрат'
)

# Данные для графика
counts = [sum(control_data["Покупки"]), sum(test_data["Покупки"])]
colors = ['#636EFA', '#EF553B']
fig = go.Figure()

# Добавление столбцов для каждой кампании
fig.add_trace(go.Bar(
    x=["Контрольная кампания"],
    y=[counts[0]],
    name="Покупки (Контрольная)",
    marker_color=colors[0],
    marker_line=dict(color='black', width=3)
))

fig.add_trace(go.Bar(
    x=["Тестовая кампания"],
    y=[counts[1]],
    name="Покупки (Тестовая)",
    marker_color=colors[1],
    marker_line=dict(color='black', width=3)
))

# Настройка макета
fig.update_layout(
    title_text='Контрольная vs Тестовая: Покупки',
    xaxis_title="Кампания",
    yaxis_title="Количество покупок",

```

```

width=800,
height=600,
showlegend=True,
legend=dict(
    title="Тип покупок",
    x=1.0,
    y=1.0,
    bgcolor='rgba(255, 255, 255, 0.5)'
),
uniformtext_minsize=12
)
# Настройка отображения значений
fig.update_traces(
    texttemplate='%{y:,.}',
    textposition='outside',
    textfont_size=16,
    textangle=0
)
fig.show()

# Проверка значимости
calculate_statistical_significance(
    control_data['Покупки'],
    test_data['Покупки'],
    'Покупки'
)

```

Приложение 3

Анализ метрик для определения конверсии

```
figure = px.scatter(data_frame=ab_data,  
                    x="Просмотры контента",  
                    y="Клики на сайте",  
                    size="Клики на сайте",  
                    color="Название кампании",  
                    trendline="ols")
```

```
figure.update_layout(  
    width=1200,  
    height=600  
)
```

```
figure.show()
```

```
figure = px.scatter(data_frame=ab_data,  
                    x="Добавлено в корзину",  
                    y="Просмотры контента",  
                    size="Добавлено в корзину",  
                    color="Название кампании",  
                    trendline="ols")
```

```
figure.update_layout(  
    width=1200,  
    height=600  
)
```

```
figure.show()
```

```
figure = px.scatter(data_frame=ab_data,  
                    x="Покупки",  
                    y="Добавлено в корзину",  
                    size="Покупки",  
                    color="Название кампании",
```

```

        trendline="ols")
figure.update_layout(
    width=1200,
    height=600
)
figure.show()
def plot_conversion_funnel(data):
    """Визуализация воронки конверсии"""
    funnel_steps = [
        "Количество показов", "Клики на сайте",
        "Добавлено в корзину", "Покупки"
    ]
    fig = make_subplots(rows=1, cols=2,
                        specs=[[{"type": "funnel"}, {"type": "funnel"}]],
                        subplot_titles=["Control Campaign", "Test Campaign"])
    colors = {"Control Campaign": "#636EFA", "Test Campaign": "#EF553B"}
    for i, campaign in enumerate(data["Название кампании"].unique()):
        campaign_data = data[data["Название кампании"] == campaign]
        values = [campaign_data[step].sum() for step in funnel_steps]
        fig.add_trace(
            go.Funnel(
                name=campaign,
                y=funnel_steps,
                x=values,
                textinfo="value+percent initial",
                marker={"color": colors[campaign]},
                textfont={"size": 14}
            ),
            row=1, col=i+1

```



```

    )
fig.update_layout(
    title_text="<b>Воронка конверсии по кампаниям</b>",
    showlegend=True,
    funnelgap=0.3,
    width=1400,
    height=600,
    margin=dict(l=100, r=100)
)
fig.show()
plot_conversion_funnel(ab_data)

```